Comput Stat (2011) 26:241–258
DOI 10.1007/s00180-010-0208-2

ORIGINAL PAPER

A primer on disease mapping and ecological regression using INLA

Birgit Schrödle · Leonhard Held

Received: 28 September 2009 / Accepted: 13 July 2010 / Published online: 13 August 2010

Abstract Spatial and spatio-temporal disease mapping models are widely used for the analysis of registry data and usually formulated in a hierarchical Bayesian framework. Explanatory variables can be included by a so-called ecological regression. It is possible to assume both a linear and a nonparametric association between disease incidence and the explanatory variable. Integrated nested Laplace approximations (INLA) can be used as a tool for Bayesian inference. INLA is a promising alternative to Markov chain Monte Carlo (MCMC) methods which provides very accurate results within short computational time. It is shown in this paper, how parameter estimates for well-known spatial and spatio-temporal models can be obtained by running INLA directly in R using the package INLA. Selected R code is shown. An emphasis is given to the inclusion of an explanatory variable. Cases of Coxiellosis among Swiss cows from 2005 to 2008 are used for illustration. The number of stillborn calves is included as time-varying covariate. Additionally, various aspects of INLA such as model choice criteria, computer time, accuracy of the results and usability of the R package are discussed.

Keywords Disease mapping · Ecological regression · INLA · Spatio-temporal models

1 Introduction

Spatial and spatio-temporal disease mapping are widespread tools for passive surveillance of a disease and, therefore, used by epidemiologists on a standard basis. The most

Electronic supplementary material The online version of this article (doi:10.1007/s00180-010-0208-2) contains supplementary material, which is available to authorized users.

B. Schrödle (✉) · L. Held
University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland
e-mail: birgit.schroedle@ifspm.uzh.ch

popular approach to spatial disease mapping was suggested in Besag et al. (1991) and developed further by several authors (Clayton and Bernardinelli 1992; Bernardinelli et al. 1995a). The methodology can be extended to the spatio-temporal case by inclusion of a linear (Bernardinelli et al. 1995b; Assunção et al. 2001) or nonparametric trend in time and time-space interactions (Knorr-Held 2000; Schmid and Held 2004). In order to investigate the association of an explanatory variable with the geographical and temporal variation in disease risk, a so-called ecological regression model can be built (Clayton et al. 1993). The effect of the covariate can be modelled in a linear or a nonparametric fashion (Fahrmeir and Lang 2001; Natario and Knorr-Held 2003).

Such spatial and spatio-temporal disease mapping models are usually formulated in a hierarchical Bayesian framework with a latent Gaussian Markov random field (GMRF) (Clayton and Bernardinelli 1992; Rue and Held 2005). So far, Markov chain Monte Carlo (MCMC) techniques have been used for Bayesian inference, but these techniques are very time-consuming since spatio-temporal disease mapping models form a complex class. Elaborate MCMC algorithms have to be used to obtain reliable posterior estimates and the MCMC output might be hard to interpret for the standard user. Integrated nested Laplace approximations (INLA) have recently been proposed as a promising alternative (Rue et al. 2009). The methodology offers very accurate approximations of the posterior marginals in short computational time. Additionally, a tool for Bayesian model choice, namely the deviance information criterion (DIC) (Spiegelhalter et al. 2002), and predictive measures as the logarithmic score (Gneiting and Raftery 2007) and the probability integral transform (PIT) (Czado et al. 2009) can be obtained.

The INLA approach is easy to apply, since a `C` program called `inla` is available. Furthermore, the `inla` program is bundled within an `R` package `INLA` to improve usability as a standard tool. As the INLA approach is a complex numerical procedure, it might still not be easy for the user to choose the right specifications and features. Hence, the `R` code needed for inference in spatial and spatio-temporal models and ecological regression using `INLA` will be introduced and the usability of the approach will be discussed.

The paper is organized as follows: First the INLA methodology is introduced briefly and possible options for the approximation algorithms are shown. In Sect. 3 some well-known spatial and spatio-temporal models for disease mapping are applied to reported Coxiellosis cases among Swiss cows from 2005 to 2008 using `INLA`. Additionally, an ecological regression analysis is performed including the number of stillborn calves as explanatory variable. In the subsequent section it is shown how the obtained output can be interpreted and used for model choice. Here the emphasis will be on spatio-temporal models. A look on computational issues and usability of the INLA approach will be taken in Sect. 4.3. A brief discussion is given in Sect. 5.

2 INLA

Spatial and spatio-temporal models as will be introduced in Sect. 3 are built as Bayesian hierarchical models with three stages: The first stage is the observational model $\pi(\mathbf{y}|\mathbf{x})$, where $\boldsymbol{y}$ denotes the observations. The vector $\boldsymbol{x}$ contains all components

of the latent Gaussian field (GMRF) $\pi(\boldsymbol{x}|\boldsymbol{\theta})$. The GRMF is typically controlled by a few hyperparameters $\boldsymbol{\theta}$, which form the third stage. Their respective prior distribution is denoted by $\pi(\boldsymbol{\theta})$. The desired posterior marginals

$$\pi(x_i|\boldsymbol{y}) = \int_{\boldsymbol{\theta}} \pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})\ \pi(\boldsymbol{\theta}|\boldsymbol{y})\ d\boldsymbol{\theta}$$

of all components of the GMRF are approximated by INLA using the finite sum

$$\widetilde{\pi}(x_i|\boldsymbol{y}) = \sum_k \widetilde{\pi}(x_i|\boldsymbol{\theta}_k, \boldsymbol{y})\ \widetilde{\pi}(\boldsymbol{\theta}_k|\boldsymbol{y})\ \Delta_k, \tag{1}$$

where $\widetilde{\pi}(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ and $\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ denote approximations of $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ and $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, respectively. This finite sum is evaluated at support points $\boldsymbol{\theta}_k$ using appropriate weights Δ_k. The $\boldsymbol{\theta}_k$'s can be obtained in two different ways, see below.

From $\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y}) = \pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y}) \times \pi(\boldsymbol{\theta}|\boldsymbol{y}) \times \pi(\boldsymbol{y})$ it follows that the posterior marginal $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ of the hyperparameters can be obtained using a Laplace approximation

$$\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}|_{\boldsymbol{x}=x^*(\boldsymbol{\theta})}$$

(Tierney and Kadane 1986), where the denominator $\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ denotes the Gaussian approximation of $\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ and $\boldsymbol{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional $\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ (Rue and Held 2005). Gaussian approximation means that the distribution of a non-normal variable is approximated by a normal distribution by matching the mode and the curvature at the mode (Rue and Held 2005, Sect. 4.4.1). According to Rue et al. (2009) it is sufficient to "numerically explore" this approximate posterior density using suitable support points $\boldsymbol{\theta}_k$ for (1). The first strategy is called GRID strategy and is computationally intensive. The mode of $\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ has to be found by some quasi-Newton method. Subsequently, the density around the mode is explored and points where the probability mass is considered as significant are selected for the integration. If the dimension h of hyperparameters included in the model is moderate ($h = 6$–12), it is computationally more efficient to use the so-called central composite design (CCD) to lay out support points in the h-dimensional space. Here, centre points are augmented with a group of star points which allow for estimating the curvature of $\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$. For more details on both methods see Rue et al. (2009). As the CCD integration scheme needs much less computational time and the differences between CCD and GRID strategy are minor, Rue et al. (2009) recommend the use of the CCD strategy for problems with high dimensionality of the hyperparameter vector $\boldsymbol{\theta}$. The difference in computer time and the resulting marginals for both strategies are briefly discussed in Sect. 4.3.

To approximate the first component of (1), namely the posterior marginal for x_i conditioned on selected values of $\boldsymbol{\theta}_k$, three different approaches are possible: A Gaussian, a full Laplace and a simplified Laplace approximation. The Gaussian approximation is fastest, but according to Rue and Martino (2007) there can be errors in the location of the posterior marginals, errors due to the lack of skewness, or both. The Gaussian approximation can be improved by using a Laplace approximation

(Tierney and Kadane 1986), but this strategy is rather time-consuming. Hence, Rue et al. (2009) introduce the so-called simplified Laplace approximation which is less expensive from a computational point of view with only a slight loss of accuracy.

To run INLA, a `C` program called `inla` is offered by the authors of Rue et al. (2009), which performs all required computations in a modular way. This program is based on the `GRMFLib`-library, which incorporates efficient algorithms for sparse matrices (Rue and Held 2005). Additionally, the computations are speeded up by the implementation of parallel computing elements. An `R`-interface called `INLA` is available to ease the usage of the `inla` program. The `inla` program is bundled within this `R` library (R Development Core Team 2005). The software can be downloaded from http://www.r-inla.org and is running in a Linux, MAC and Windows environment. For the analyses within this paper we used the `INLA` library built on the 28th of April 2010. The respective `R` code is shown where it was considered as helpful. The data and further `R` code can be found within Online Resource 1.

3 Review of spatial and spatio-temporal disease mapping models and their specification using `INLA`

The following sections give an introduction to the data and show the specification of selected spatial and spatio-temporal disease mapping models using `INLA`.

3.1 Data—Cases of Coxiellosis among cows in Switzerland, 2005–2008

Coxiellosis is a widespread infectious, endemic disease caused by the bacterium *coxiella burnetii* among ruminant animals (Aitken 1989). In most cases it is subclinical, but it can be the reason for an abortion in a late phase of the pregnancy or a stillbirth (Woldehiwet 2004). The spread of the bacterium can take place through ticks, but happens as well from animal to animal by airborne infection as the bacterium is present in abortion products and excreted by diseased animals in their milk, urine and excrement. Special attention must be payed to this disease as it is a so-called zoonosis that can also affect humans (Q fever); such epidemics have been observed in Switzerland (Dupuis et al. 1987).

The data considered are cases of Coxiellosis among cows reported to the Swiss Federal Veterinary Office from 2005 to 2008. A herd is marked as infected, if one or more diseased cows were detected. The number of cases is available on a yearly basis for 184 regions of Switzerland. Additionally, data from the Principality of Liechtenstein is included. As shown in Table 1, the number of reported cases has constantly been rising during the last four years. Hence, it is of interest if a significant rise in reported cases took place and what the spatial distribution of the disease within Switzerland looks like.

Table 1 Number of reported cases of Coxiellosis in cows per year, 2005–2008

Year	2005	2006	2007	2008
n	30	45	54	61

As Coxiellosis is a widespread disease, it is obvious from Table 1 that massive underreporting must be present, although the disease is notifiable in Switzerland. Switzerland is a confederation of 26 cantons, which consist of one or more regions. The cantonal veterinary authorities are responsible for the realization of federal veterinary legislation in each affiliated region. In Schrödle et al. (2010) it was found that the number of reported cases within one region might depend on the canton it belongs to. Hence, cantons are considered as a second, coarser spatial grid.

In Sect. 3.2 the cases from 2008 only are used as response variable for spatial disease mapping, while the spatio-temporal disease mapping in Sect. 3.3 is illustrated using all cases from 2005 to 2008.

3.2 Spatial disease mapping

Under the rare disease assumption it is usually assumed that the number of disease cases y_i in region $i = 1, \ldots, 185$ is Poisson distributed with parameter λ_i, which can be interpreted as the relative risk of the disease in the respective region. Additionally, the number of herds m_i is included in the model as an offset to adjust for the different number of herds at risk. In the standard formulation established by Besag et al. (1991) the relative risk parameter λ_i is specified as

$$\eta_i = \log(\lambda_i) = \log(m_i) + \mu + v_i + \psi_i. \tag{2}$$

This model will be called BYM1. It contains a spatially unstructured component v_i (variable name (vn): `region.nu`) which is i.i.d. normally distributed with zero mean and unknown precision τ_v, whereas ψ_i (vn: `region.psi`) is assumed to be structured in space. To account for the assumption that geographically close areas have similar incidence rates the spatially structured component ψ_i is modelled as an intrinsic Gaussian Markov random field (IGMRF) with unknown precision τ_ψ (Rue and Held 2005). This specification is also called a conditionally autoregressive (CAR) prior (Banerjee et al. 2004). To ensure identifiability of the intercept μ a sum-to-zero constraint must be imposed on the ψ_i's. The variables `region.nu` and `region.psi` are identical, but two different objects have to be specified within INLA.

As discussed in Schrödle et al. (2010), the number of reported cases per region might depend on the canton a region belongs to. For investigation of this fact, (2) is extended to a second, coarser spatial level: An i.i.d. random effect α_j (vn: `canton.alpha`) for each of the 26 cantons of Switzerland and Liechtenstein is added ($j = 1, \ldots, 27$). The resulting linear predictor is

$$\eta_i = \log(\lambda_i) = \log(m_i) + \mu + v_i + \psi_i + \alpha_{j(i)}. \tag{3}$$

The extended model will be called BYM2.

The choice of hyperpriors for disease mapping models is discussed in Bernardinelli et al. (1995a). As proposed we use Prior B, Ga(1, 0.01) (vn: `prior.nu`), as hyperprior for τ_v and τ_α. The prior for τ_ψ was adjusted for the structure of the Swiss graph and chosen as Ga(1, 0.018) (vn: `prior.psi`) (Bernardinelli et al. 1995a).

To run these models in INLA, the linear predictor of the model has to be specified as a formula object in R using the function f() for smooth effects. Subsequently, the specified model can be run using inla().

The type of a smooth effect can be specified in f() using, e.g., model="iid" for an i.i.d. random effect and "besag" for an IGMRF like ψ. The respective graph-file (e.g. "switzerland.graph") containing the neighbourhood structure has to be specified as well. The hyperpriors for the precision parameters of the smooth effects (argument: param) have to be chosen and linear constraints can be set (argument: constraint). For the "besag"-prior a sum-to-zero constraint is imposed as default. Within the inla()-call further options for the INLA algorithm can be set. Here it can, e.g., be specified if quantities for predictive measures (cpo=1) and the DIC (dic=1) should be computed and which strategy for the approximation of the latent Gaussian field and the posterior marginals of the hyperparameters $\boldsymbol{\theta}$ should be used. The default choice is the simplified Laplace approximation (SLP) and the CCD strategy. As shown in Held et al. (2010), the accuracy of the SLP approximation is often not sufficient for the computation of predictive measures. Hence, the full Laplace approximation was chosen in the following application (strategy="laplace"). A dataframe can be specified using the argument data; the offset (vn: offset) for a Poisson model is given to INLA via E. The vector Y.cox of the dataset cox.08 contains the number of Coxiellosis cases per region in 2008. For more details see the inla manual (Martino and Rue 2009).

The resulting model specification and the call to fit model BYM1 is

```
> f.BYM1<-Y.cox~f(region.nu,model="iid",param=prior.nu)+
+               +f(region.psi,model="besag",param=prior.psi,
+                  graph.file="switzerland.graph")
> BYM1<-inla(f.BYM1,family="poisson",E=offset,data=cox.08,
+            control.inla=list(strategy="laplace"),
+            control.compute=list(dic=1,cpo=1))
```

Using names(BYM1) the components of the output can be seen. For example,

```
> round(BYM1$summary.random$region.nu[1, ], digits = 4)
```

```
  ID   mean     sd 0.025quant 0.5quant 0.975quant    kld
1  0 -0.283 1.0572    -2.5108  -0.2372     1.6483 0.0028
```

returns the results for the unstructured effect of region 1. Standard outputs are the posterior mean, standard deviation, 2.5%-, 50%- and 97.5%-quantiles and the symmetric Kullback-Leibler distance (SKLD) between the Gaussian and the simplified/full Laplace approximation, which is derived from the Kullback-Leibler discrepancy (KLD). The KLD is a measure to quantify the divergence between two density functions, but it is not symmetric (Kullback and Leibler 1951). To solve this problem the SKLD is defined as the sum of the KLD's measured in both directions (Wood and Kohn 1998; Moreno et al. 2004). Model BYM2 can be specified in a similar fashion; the respective results are discussed in Sect. 4.1.

3.3 Spatio-temporal disease mapping

To find out, if there has been a statistically significant linear rise in reported cases of Coxiellosis from 2005 to 2008, a spatio-temporal disease mapping model is adopted in the following section. This model is analogous to Bernardinelli et al. (1995b), but expanded by a cantonal effect α_j as case reporting might be biased with regard to the cantonal affiliation of a region. The linear predictor can be written as

$$\eta_{it} = \log(m_i) + \mu + v_i + \psi_i + \alpha_{j(i)} + (\beta_1 + \delta_{j(i)}) \cdot t. \tag{4}$$

This model includes the same components as (3), but a main linear time trend β_1 (vn: `time.beta1`) and a so-called differential trend δ_j (vn: `differential.delta`) for each canton are added. The effect δ_j is modelled as a random slope and accounts for cantonal departures from the main linear time trend. As it is necessary to allow for correlation between intercept and slope in a random slope model, it is assumed that $(\alpha_j, \delta_j)^T$ follows a bivariate normal distribution with zero mean and some unknown precision matrix, to which a Wishart prior is assigned (Bernardinelli et al. 1995b). Using INLA $(\alpha_j, \delta_j)^T$ can be defined using two components `model="2diidwishartpart0"` and `"2diidwishartpart1"`, respectively. Four parameters have to be specified within the `"2diidwishartpart0"`-component. In this application these parameters are chosen as `prior.wishart=c(4,1,1,0)`; this choice was used in Schrödle et al. (2010) for a similar setting and checked for sensitivity. For the differential trend δ_j appropriate weights (vn: `time`) given by the timepoints have to be introduced additionally.

This model is called ST1 and the model formula is defined as

```
> f.ST1<-Y.cox~f(region.nu,model="iid",param=prior.nu)+
+              +f(region.psi,model="besag",param=prior.psi,
+                 graph.file="switzerland.graph")+
+              +f(canton.alpha,model="2diidwishartpart0",
+                 param=prior.wishart)+
+              +f(differential.delta,time,
+                 model="2diidwishartpart1")+
+              +time.beta1
> ST1<-inla(f.ST1,family="poisson",E=offset,data=cox,
+            control.inla=list(strategy="laplace"),
+            control.compute=list(dic=1,cpo=1))
```

The vectors `Y.cox`, `offset`, `region.psi`, `region.nu` and `canton.alpha` in the dataset `cox` are four times as long as the corresponding vectors in the dataset `cox.08` from Sect. 3.2, since four years are included in the spatio-temporal analysis. The vectors `canton.alpha` and `differential.delta` are identical, but two different variables must be specified within INLA.

Another option would be to assume a nonparametric trend in time as proposed in Knorr-Held (2000). This approach was not considered here as only a short time interval is taken into account, but its implementation in INLA is possible (Schrödle et al. 2010).

3.4 Ecological regression

The model in the preceding section can be extended to an explanatory variable to investigate its association with the geographical and temporal variation in disease risk (Clayton and Bernardinelli 1992; Clayton et al. 1993).

As noted in Sect. 3.2, case reporting in Switzerland might be biased by factors like, e.g., temporally varying disease awareness. So it is of interest, if the rise in reported cases (see Table 1) can be accounted to a "real" rise of disease incidence. Since Coxiellosis can cause the stillbirth of a calf (Aitken 1989), a spatial and temporal association between the number of stillborn calves and reported Coxiellosis cases within one region would indicate a "real" rise in the incidence of the disease. The number of stillborn calves is available for each region and year and has constantly been growing since 2005, see Table 2. The covariate was square-root transformed before the analysis (vn: `ncalves.beta2`); a boxplot of the respective values can be found in Fig. 1.

In Clayton and Bernardinelli (1992) and Clayton et al. (1993) it has been suggested to assume a linear relationship for the explanatory variable. Hence, (4) is expanded by inclusion of a linear covariate z_{it}

$$\eta_{it} = \log(m_i) + \mu + \nu_i + \psi_i + \alpha_{j(i)} + (\beta_1 + \delta_{j(i)}) \cdot t + \beta_2 \cdot z_{it}. \tag{5}$$

This model will be denoted as model ST2.

Natario and Knorr-Held (2003) have proposed to replace the linear effect of z_{it} with a smooth nonparametric function f_z. The resulting model can be written as

$$\eta_{it} = \log(m_i) + \mu + \nu_i + \psi_i + \alpha_{j(i)} + (\beta_1 + \delta_{j(i)}) \cdot t + f_z(z_{it}). \tag{6}$$

This model will be called ST3.

Table 2 Number of stillborn calves per year

Year	2005	2006	2007	2008
n	15326	23044	25289	26911

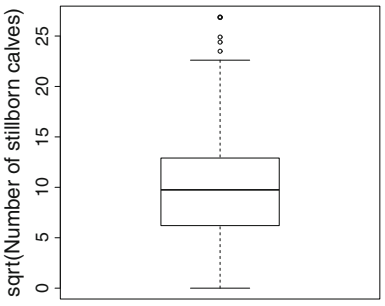

Fig. 1 Boxplot of the square root of the number of stillborn calves, 2005–2008

In the easy case the covariate z_{it} can take only K equally spaced levels $g_1 < \cdots < g_k < \cdots < g_K$. Then $\gamma_k = f_z(z_{it} = g_k)$ is assumed to follow a random walk of second order on regular locations with joint density

$$\pi(\boldsymbol{\gamma}|\tau_\gamma) \propto \exp\left(-\frac{\tau_\gamma}{2}\sum_{k=2}^{K-1}(\gamma_{k+1} - 2\gamma_k + \gamma_{k-1})^2\right) \tag{7}$$

and precision τ_γ (Fahrmeir and Lang 2001; Rue and Held 2005). This is a natural assumption, as a random walk of second order models deviations from a linear trend (Natario and Knorr-Held 2003). In addition, it is appropriate for representing smooth curves with a small curvature and is computationally convenient due to its Markov property. A sum-to-zero constraint has to be imposed on the γ_k's to ensure identifiability of μ. As the levels of the covariate in this application are not equally spaced and the use of equal spaces would increase the dimension of the model (Lindgren and Rue 2008), (7) has to be extended to the more general case with non-equally spaced levels. In Fahrmeir and Lang (2001) it has been suggested to include appropriate weights, but this approach leads to inconsistencies regarding variances for the case of non-equally spaced levels (Lindgren and Rue 2008). Hence, a new approach has been proposed in Lindgren and Rue (2008), where (7) is interpreted as an approximated Galerkin solution to the stochastic differential equation $f''(t) = \mathrm{d}W(t)/\mathrm{d}t$, where $W(t)$ is a Wiener process (Rue and Held 2005). This approach does not show inconsistencies regarding the variances and its covariance properties converge to those of a continuous RW2 process as the grid of the observed levels gets more dense. It is computationally convenient with negligible errors and, hence, implemented in INLA when using option `model="rw2"`. Therefore, model (6) can be specified using

```
> f.ST3<-Y.cox~f(region.nu,model="iid",param=prior.nu)+
+              +f(region.psi,model="besag",param=prior.psi,
+                 graph.file="switzerland.graph")+
+              +f(canton.alpha,model="2diidwishartpart0",
+                 param=prior.wishart)+
+              +f(differential.delta,time,
+                 model="2diidwishartpart1")+
+              +time.beta1+
+              +f(ncalves.gamma,model="rw2",param=prior.gamma)
```

Within INLA it is also possible to model $\boldsymbol{\gamma}$ as a continuous time random walk of second order (Rue and Held 2005). This approach might be more time-consuming compared to the discretized approach in Lindgren and Rue (2008). To run it, the option `"rw2"` has to be replaced by `"crw2"`.

Care has to be taken concerning the prior chosen for the variance $\sigma_\gamma^2 = 1/\tau_\gamma$, as its interpretation depends on the levels taken by the covariate and the distance ξ between successive values of g_k. As noted in Berzuini and Clayton (1994), the ratio of the prior mode of σ_γ, which specifies the prior belief in smoothness, and the squared distance ξ^2 should be kept constant when varying the parameters of the hyperprior. Usually an inverse gamma distribution IGa(a,b) with prior mode $b/(a+1)$ is adopted as hyperprior for σ_γ^2. In Natario and Knorr-Held (2003) it is recommended to use an IGa(1, 0.00005) prior (vn: `prior.gamma`) for non-equally spaced covariates with an average distance of 1. Hence, the values of the covariate (vn: `ncalves.gamma`)

in this application were scaled in a way that this requirement is satisfied. In Natario and Knorr-Held (2003) it was found that the nonparametric trend is sensitive to the choice of the prior; for the application at hand this will be investigated in more detail in Sect. 4.

4 Results and model choice

The following sections show, how the INLA output for all models specified in Sect. 3 can be used for model choice and interpretation. To shorten considerations, Sect. 4.1 deals only with model choice whereas in Sect. 4.2 results for the spatio-temporal analysis are presented as well. Some issues with regard to computer time and the use of different approximation techniques are briefly discussed in Sect. 4.3.

4.1 Spatial disease mapping—Model choice

As noted before, several quantities for model choice and model calibration are available by INLA. In order to decide which model provides the best trade-off between model fit and complexity, the DIC is given as a well-known Bayesian model choice criterion (Spiegelhalter et al. 2002). The conditional predictive ordinates (CPO's) which facilitate the computation of the cross-validated logarithmic score for model choice (Gneiting and Raftery 2007) are given as well as the probability integral transform (PIT), which can be used to assess calibration of out-of-sample predictions (Czado et al. 2009). The use of these measures is exemplary shown for model choice between BYM1 and BYM2.

The DIC is the sum of a measure of model fit, the posterior mean of the deviance $\bar{D}$, and model complexity, the effective number of parameters p_D, and is addressed using

```
> BYM1$dic

                                     [,1]
mean of the deviance           158.37680
deviance of the mean           125.12526
effective number of parameters  33.25154
dic                            191.62834
```

```
> BYM2$dic

                                     [,1]
mean of the deviance           158.27891
deviance of the mean           130.09263
effective number of parameters  28.18627
dic                            186.46518
```

The smaller the DIC, the better the trade-off between model fit and complexity. The posterior deviance and the number of effective parameters in model BYM1 are slightly larger than in model BYM2. Hence, the DIC value of model BYM2 is smaller. The logarithmic score (Gneiting and Raftery 2007) can be computed as

```
> lsBYM1 <- -mean(log(BYM1$cpo))
> lsBYM2 <- -mean(log(BYM2$cpo))
> round(lsBYM1, digits = 3)

[1] 0.55

> round(lsBYM2, digits = 3)

[1] 0.532
```

The smaller the resulting score, the better the predictive quality of the model. As the score for model BYM2 is a bit smaller the predictive quality for BYM2 is better. The calibration of both models can be checked by plotting an adjusted PIT histogram as suggested by Czado et al. (2009) using the values provided in, for example, `BYM1$pit`. The results can be seen in Fig. 2; the histograms are close to uniform. Both models are almost perfectly calibrated, but the calibration of model BYM2 seems to be slightly better. Hence, BYM2 is preferred.

4.2 Spatio-temporal disease mapping—Results and model choice

The results for model ST1 are called using

```
> round(ST1$summary.fixed, digits = 4)

               mean     sd 0.025quant 0.5quant 0.975quant    kld
(Intercept) -7.6492 0.2780    -8.2219  -7.6395    -7.1272 0.5287
time.beta1   0.3400 0.1548     0.0514   0.3342     0.6637 0.0573

> round(ST1$summary.hyperpar[, c(1, 2)], digits = 4)

                                                    mean      sd
Precision for region.nu                           3.9539  1.7894
Precision for region.psi                         61.7972 45.6382
Precision for canton.alpha (first component)      1.0040  0.3966
Precision for canton.alpha (second component)     5.3926  2.2544
Rho for canton.alpha                             -0.4379  0.2282
```

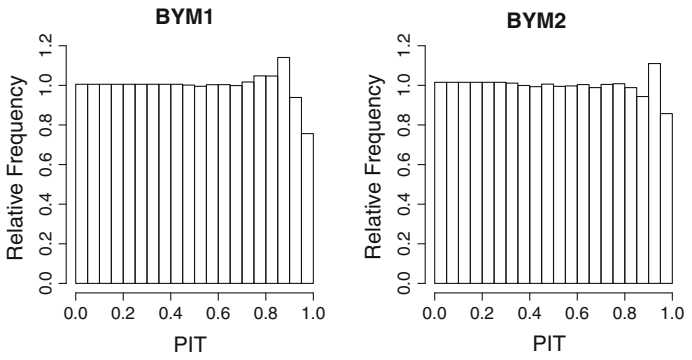

Fig. 2 Adjusted PIT-histograms for models BYM 1 and BYM 2

A significantly positive linear time trend can be observed. Hence, a significant rise in disease incidence has taken place over the last four years. The summary of the obtained posterior estimates for the hyperparameters (τ_v, τ_ψ, τ_α, τ_δ, ρ) shows, that the spatially structured regional effect can almost be neglected. Spatially structured heterogeneity is covered on a coarser resolution by the cantonal trend. The estimated correlation ρ between the cantonal and the differential trend is negative. So, the higher the cantonal intercept the less steep than the main time trend is the time trend of the respective canton. This fact can also be seen in Fig. 3a. It shows the individual time trend for each canton ($\mu + \alpha_j + (\beta_1 + \delta_j) \cdot t$). Cantons with a time trend that is significantly different from the main time trend are plotted with various line types. The two cantons with the highest disease incidence, namely Jura and Obwalden, show a significantly negative time trend, while it is positive for almost all other cantons. A plot of the mean spatial incidence of Coxiellosis for the years 2005 to 2008 is shown in Fig. 3b.

To assess the significance of the explanatory variable the output for model ST2 has to be considered. The results for the fixed effects can be called using

```
> round(ST2$summary.fixed, digits = 3)
```

	mean	sd	0.025quant	0.5quant	0.975quant	kld
(Intercept)	-9.232	0.550	-10.359	-9.216	-8.196	0.283
time.beta1	0.212	0.161	-0.091	0.207	0.545	0.036
ncalves.beta2	0.124	0.034	0.059	0.123	0.192	0.050

The number of stillborn calves is significantly positive associated with the incidence of Coxiellosis within one region. This indicates that a real rise in disease incidence has taken place. To assure that the significance of the covariate is not confounded with the positive temporal trend, an ecological analysis has been conducted for each year separately. A significant association was found for each year, except for 2007.

The influence of the number of stillborn calves was modelled in a nonparametric fashion in models ST3 (`"rw2"`) and ST4 (`"crw2"`). Results for model ST3 using

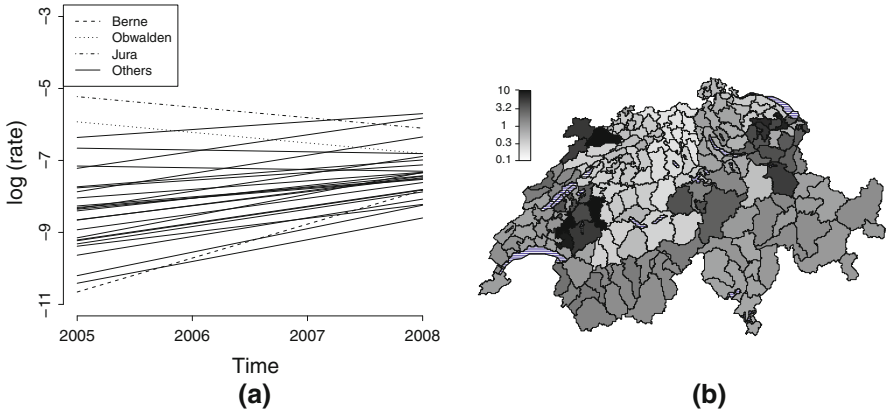

Fig. 3 **a** Linear time trend for each canton, cantons with a significantly different time trend are plotted using various line types (ST1); **b** Relative incidence for Coxiellosis, 2005–2008 (ST1)

the prior suggested in Natario and Knorr-Held (2003) are shown in the first plot of Fig. 4, including a 95%-confidence interval. To check, if the estimated linear effect is contained in the confidence interval of the nonparametric effect, it is also plotted. The larger (pointwise) confidence intervals for more extreme values are a typical feature of nonparametric smoothing methods. The results for `"rw2"` and `"crw2"` are almost identical, except for negligible differences in the tails of the curves.

Figure 4 also shows the results of a sensitivity analysis regarding the IGa(a,b)-hyperprior on the variance $\sigma_\gamma^2 = 1/\tau_\gamma$ for choices where the first parameter, the so-called shape parameter, is kept constant, but the prior mode $b/(a+1)$ increases from left to right. As noted by Natario and Knorr-Held (2003) the resulting curve is highly sensitive to this choice. The larger the prior mode, the more wiggly is the curve. This is the case for the `"rw2"` as well as for the `"crw2"` specification. Additionally it was found that the results barely change, if the prior mode is kept constant and only the shape parameter is varied.

Regarding model choice the same quantities as in Sect. 4.1 can be considered. Table 3 shows the results for the DIC and the logarithmic score. The DIC for all models including the covariate is smaller than for the model without covariate which, again, suggests a significant association between the number of stillborn calves and the Coxiellosis incidence. Not only the posterior deviance, but also the number of effective parameters is smaller when comparing models ST1 and ST2. Model ST3 provides a better fit than model ST2, but, as expected, the number of effective parameters is

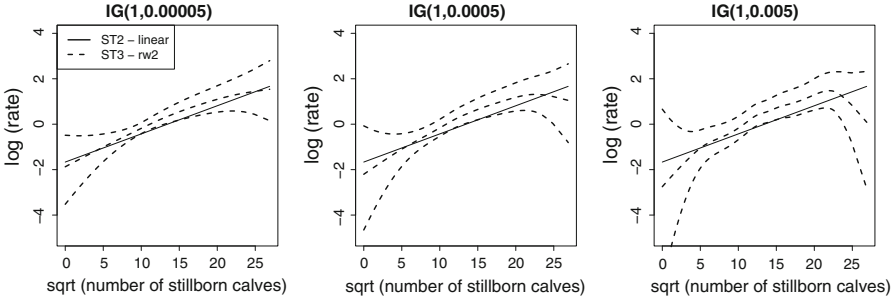

Fig. 4 Estimated nonparametric trend γ for model ST3 (*dashed line*); additionally to the estimated posterior mean a pointwise 95%-confidence interval is plotted. The estimated linear trend (ST2) is plotted as well (*solid line*). The results are shown for different specifications of the prior fors σ_γ^2

Table 3 Summary of the posterior mean of the deviance ($\overline{D}$), the number of effective parameters (p_D) and the resulting sum, the DIC, as a measure of trade-off between model fit and complexity for models ST1, ST2, ST3 and ST4; additionally, the logarithmic score ($\overline{\text{LS}}$) is given

Model	$\overline{D}$	p_D	DIC	$\overline{\text{LS}}$
ST1	603.7	48.6	652.3	0.455
ST2	600.5	43.5	643.9	0.448
ST3	599.2	44.1	643.3	0.449
ST4	599.7	43.9	643.6	0.449

slightly higher. The best trade-off between model fit and complexity is offered by model ST3. The mean logarithmic score is smallest for models ST2. The PIT histograms are not shown, but close to uniform; hence, all models are well calibrated.

As a general result it can be derived that a significant rise in reported cases has taken place. There is a positive association between the number of stillborn calves and the disease incidence within one region. A linear relationship might be sufficient to model this association. A drawback concerning a nonparametric formulation of the covariate is the high sensitivity towards the choice of the hyperprior.

4.3 Some comments on computer time and the accuracy of approximations

As noted in Sect. 2, two strategies for the exploration of the posterior marginal $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ exist, namely the GRID and the CCD strategy. Using `INLA` these strategies can be chosen using the options `int.strategy="grid"` and `"CCD"`, respectively. The CCD strategy is less precise, but takes much less computational time, as fewer support points for the integration in (1) are needed. As an example consider a spatio-temporal model like (4), which contains $h = 5$ hyperparameters $(\tau_v, \tau_\psi, \tau_\alpha, \tau_\delta, \rho)$. Using the `INLA` default configurations for the density of the grid, the GRID strategy needs $5^h = 5^5 = 3125$ support points, while only 27 are needed for the CCD strategy. The more hyperparameters are included in the model, the larger the difference in the number of support points between the GRID and the CCD strategy. A second issue is the chosen approximation for the latent Gaussian field $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$. Possible strategies are `strategy="gaussian"`, `"simplified.laplace"` and `"laplace"`, as described in Sect. 2.

The resulting computer time for model ST1 for all configurations on a Laptop with Intel(R) Core(TM) 2 Duo CPU T9300 2.50 GHz processor is summarized in Table 4. The computer time needed for the GRID strategy is much higher than the time needed for the CCD strategy. The computer time increases as well when switching from the Gaussian to a simplified Laplace and a full Laplace approximation, respectively. The switch from the Gaussian to the simplified Laplace approximation takes less time than the switch to the full Laplace approximation.

The function `inla.hyperpar()` was applied to the CCD and GRID strategy results to obtained more precise approximations of the posterior marginals of single hyperparameters. The resulting CCD and GRID curves are identical for each

Table 4 First line: The computer time the `R` user has to wait for a result of model ST1 (in seconds); the model was run using the CCD and GRID strategy for approximation of $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ and all three approximation techniques for $\pi(x_i|\boldsymbol{y}, \boldsymbol{\theta})$. Second line: Number of observations where the computation of the predictive quantities is problematic or unreliable (in brackets)

	Gauss		SLP		FL	
	CCD	GRID	CCD	GRID	CCD	GRID
Computer time	24.73	317.45	28.65	672.74	171.45	13310.86
# of failures	164 (9)	174 (0)	22 (1)	25 (0)	0 (0)	0 (0)

hyperparameter. Hence, the CCD strategy is sufficient for these data. In Figure 5 the posterior marginals for all hyperparameters of model ST1 resulting from the CCD strategy are shown on log scale (except for ρ).

With regard to a comparison of the different approximation methods, the components of the latent field with the largest discrepancy between the possible approximations were determined. This was done for all random effects (i.e. $\boldsymbol{\nu}$, $\boldsymbol{\psi}$, $\boldsymbol{\alpha}$, $\boldsymbol{\delta}$) using the maximum symmetric Kullback-Leibler distance (SKLD) between the Gaussian and the full Laplace approximation. The resulting plots are shown in Fig. 6. For $\boldsymbol{\nu}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ a shift in location can be detected for the Gaussian approximation. The results for the simplified and the full Laplace approximation are virtually identical. Hence, the simplified Laplace approximation gives satisfactory results in terms of accuracy.

Regarding the predictive measures given by INLA the simplified Laplace approximation might not be sufficient though. As already noted in Sect. 3.2 and derived in Held et al. (2010), the approximation of the predictive measures as shown in Rue et al. (2009) might fail, if the approximation of the latent field is not accurate enough. This is due to an insufficient exploration of the tail properties of involved densities. Hence, the full Laplace approximation might be obligatory to get reliable results. A feature of INLA is that it outputs a file which indicates the observations, where computation of the predictive measures failed. This file can be addressed using, for example,

```
> ST1$failure
```

It contains a flag for each observation. If the flag is 0, the computation of the predictive measures cpo and pit was not problematic. If it is larger than 0, there

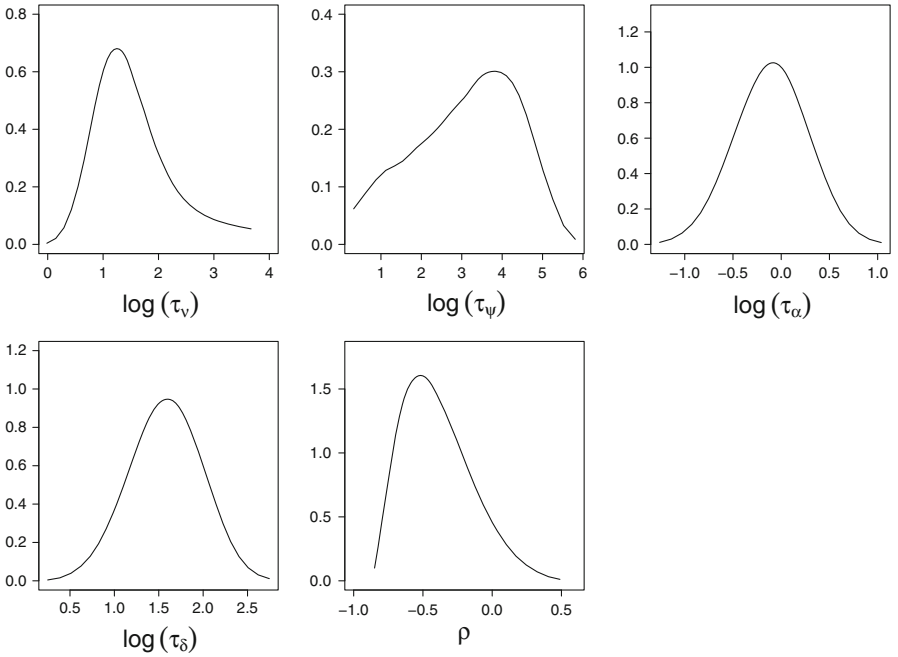

Fig. 5 Posterior marginals of hyperparameters included in model ST1 on log scale (except for ρ), estimated using the CCD strategy. The results for the CCD and GRID strategy are identical

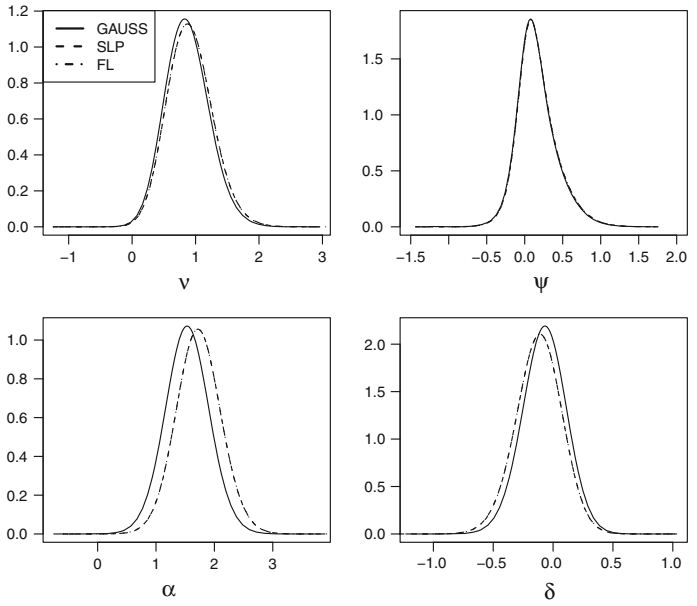

Fig. 6 Posterior marginals for effects included in model ST1, estimated using the Gaussian (*solid line*), simplified Laplace (*dashed*) and full Laplace (*dashed* and *dotted*) approximation; results are shown for the marginals with the maximum symmetric Kullback-Leibler distance for the respective effect

were some problems; if it is equal to 1, the obtained results are considered to be unreliable (Martino and Rue 2009). In general, results with a `failure`-flag larger than 0 should not be used. The number of observations with problematic/unreliable results for each strategy for model ST1 is shown in Table 4. The problem is mainly solved by using the full Laplace approximation. If this is not the case, the respective measures have to be computed manually by leaving out one observation in turn, re-running INLA and computing the leave-one-out predictive distribution from the respective `INLA` output (Held et al. 2010).

5 Discussion

As shown within this paper, INLA can be used for Bayesian inference in spatial and spatio-temporal disease mapping models. Additionally, ecological regression can be performed involving a linear or nonparametric association between an explanatory variable and the disease incidence. The available `R` interface `INLA` can easily be handled by the user and the obtained results are useful for interpretation and suitable for model choice using the DIC or predictive measures like the logarithmic score and the PIT histogram. As INLA is a numerical approach and has a complex nature, different options for the exploration algorithm of the posterior marginals of the hyperparameters and approximation methods for the latent Gaussian field are available. This fact might make the first steps with `INLA` difficult for the standard user. As noted in Sect. 4.3, the default strategies give satisfactory results in this application.

The computation of predictive measures often requires the use of the full Laplace approximation to obtain reliable results.

Acknowledgments Financial support by the Swiss Federal Veterinary Office (BVET) and the Swiss National Science Foundation (SNF) is gratefully acknowledged.

References

Aitken I (1989) Clinical aspects and prevention of Q fever in animals. Eur J Epidemiol 5(4):420–424
Assunção R, Reis I, Oliveira C (2001) Diffusion and prediction of leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model. Stat Med 20(15):2319–2335
Banerjee S, Carlin B, Gelfand A (2004) Hierarchical modeling and analysis for spatial data. Chapman & Hall/CRC, London
Bernardinelli L, Clayton D, Montomoli C (1995) Bayesian estimates of disease maps: how important are priors? Stat Med 14:2411–2431
Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M (1995) Bayesian analysis of space-time variation in disease risk. Stat Med 14:2433–2443
Berzuini C, Clayton D (1994) Bayesian analysis of survival on multiple time scales. Stat Med 13:823–838
Besag J, York J, Mollié A (1991) Bayesian image restoration with two applications in spatial statistics. Ann Inst Stat Math 43(1):1–59
Clayton D, Bernardinelli L (1992) Bayesian methods for mapping disease risk. In: Cuzick J (ed) et al Geographical and environmental epidemiology. Methods for small area studies. Oxford University Press, Oxford, pp 205–220
Clayton D, Bernardinelli L, Montomoli C (1993) Spatial correlation in ecological analysis. Int J Epidemiol 22(6):1193–1202
Czado C, Gneiting T, Held L (2009) Predictive model assessment for count data. Biometrics 65(4):1254–1261
Dupuis G, Petite J, Péter O, Vouilloz M (1987) An important outbreak of human Q fever in a Swiss alpine valley. Int J Epidemiol 16(2):282–287
Fahrmeir L, Lang S (2001) Bayesian inference for generalized additive mixed models based on Markov random field priors. J R Stat Soc Ser C 50(2):201–220
Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 102(477):359–378
Held L, Schrödle B, Rue H (2010) Posterior and cross-validatory predictive checks: a comparison of MCMC and INLA. In: Kneib T, Tutz G (eds) Statistical modelling and regression structures—Festschrift in honour of Ludwig Fahrmeir. Physica-Verlag, Heidelberg
Knorr-Held L (2000) Bayesian modelling of inseparable space-time variation in disease risk. Stat Med 19:2555–2567
Kullback S, Leibler R (1951) On information and sufficiency. Ann Math Stat 22(1):79–86
Lindgren F, Rue H (2008) On the second-order random walk model for irregular locations. Scand J Stat 35:691–700
Martino S, Rue H (2009) Implementing approximate Bayesian inference using Integrated Nested Laplace Approximation: A manual for the inla program. Technical report, Norwegian University of Science and Technology Trondheim
Moreno P, Ho P, Vasconcelos N (2004) A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications. In: Thrun S, Saul L, Schölkopf B (eds) Advances in neural information processing systems 16. MIT Press, Cambridge
Natario I, Knorr-Held L (2003) Non-parametric ecological regression and spatial variation. Biom J 45(6):670–688
R Development Core Team (2005) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org, ISBN 3-900051-07-0
Rue H, Held L (2005) Gaussian Markov random fields. Chapman & Hall/CRC, London
Rue H, Martino S (2007) Approximate Bayesian inference for hierarchical Gaussian Markov random field models. J Stat Plan Inference 137:3177–3192

Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). J R Stat Soc Ser B 71:319–392
Schmid V, Held L (2004) Bayesian extrapolation of space-time trends in cancer registry data. Biometrics 60:1034–1042
Schrödle B, Held L, Riebler A, Danuser J (2010) Using INLA for the evaluation of veterinary surveillance data from Switzerland: a case study. Technical report, University of Zurich
Spiegelhalter D, Best N, Carlin B, van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). J R Stat Soc Ser B 64(4):583–639
Tierney L, Kadane JB (1986) Accurate approximations for posterior moments and marginal densities. J Am Stat Assoc 81(393):82–86
Woldehiwet Z (2004) Q fever (Coxiellosis): epidemiology and pathogenesis. Res Vet Sci 77(2):93–100
Wood S, Kohn R (1998) A Bayesian approach to robust binary nonparametric regression. J Am Stat Assoc 93(441):203–213